

2020

## Data Mining Raman Microspectroscopic Responses of Cells to Drugs in Vitro Using Multivariate Curve Resolution-Alternating Least Squares

David Perezguaita

*Technological University Dublin, david.perezguaita@tudublin.ie*

Guillermo Quintas

*Leitat Technological Centre, Barcelona, Spain*

Zeineb Farhane

*Technological University Dublin, zeinab.farhane@mytudublin.ie*

Roma Tauler

*IDAEA-CSIC, Barcelona, Spain*

Hugh J. Byrne

Follow this and additional works at: <https://arrow.tudublin.ie/nanolart>  
*Technological University Dublin, hugh.byrne@tudublin.ie*



Part of the [Pharmacology, Toxicology and Environmental Health Commons](#)

### Recommended Citation

Perez-Guaita, D., Quintas, G., Farhane, Z., Tauler, R. & Byrne, H.J. (2020). Data mining Raman microspectroscopic responses of cells to drugs in vitro using multivariate curve resolution-alternating least squares. *Talanta*, 208, 120386. doi:10.1016/j.talanta.2019.120386

This Article is brought to you for free and open access by the NanoLab at ARROW@TU Dublin. It has been accepted for inclusion in Articles by an authorized administrator of ARROW@TU Dublin. For more information, please contact [yvonne.desmond@tudublin.ie](mailto:yvonne.desmond@tudublin.ie), [arrow.admin@tudublin.ie](mailto:arrow.admin@tudublin.ie), [brian.widdis@tudublin.ie](mailto:brian.widdis@tudublin.ie).



This work is licensed under a [Creative Commons Attribution-NonCommercial-Share Alike 3.0 License](#)

# Data mining Raman Microspectroscopic Responses of Cells to Drugs in Vitro using Multivariate Curve Resolution-Alternating Least Squares

David Perez-Guaita<sup>1\*</sup>, Guillermo Quintas<sup>2,3</sup>, Zeineb Farhane<sup>1</sup>, Roma Tauler<sup>4</sup> and Hugh J. Byrne<sup>1</sup>

<sup>1</sup> FOCAS Research Institute, Technological University Dublin, City Campus, Dublin 8, Ireland

<sup>2</sup> Health and Biomedicine, Leitat Technological Centre, Barcelona, Spain

<sup>3</sup> Unidad Analítica, Health Research Institute La Fe, Valencia, Spain

<sup>4</sup> IDAEA-CSIC, Barcelona, Spain

\* david.perezguaita@dit.ie

**Keywords: Raman Spectroscopy; Drug Uptake; In vivo cell analysis, Chemometrics, MCR-ALS**

## Abstract

Raman microspectroscopy is gaining popularity for the analysis of time-dependent biological processes such as drug uptake and cellular response. It is a label-free technique which acquires signals from a large variety of components, including cell biomolecules and exogenous compounds such as drugs and nanoparticles, and is commonly employed for *in vitro* analysis of cells and cell populations with no labelling or staining required. By monitoring the changes to the Raman spectra of the cell as a result of a perturbing agent (e.g. inoculation of a drug or toxic agent), one can study the associated changes in cell biochemistry involved in both, the disruption and the subsequent cellular response. The main challenge is that the Raman spectra should be data mined in order to extract the information corresponding to the different actors involved on the process. Here, we study the application of multivariate curve resolution-alternating least squares (MCR-ALS) for extracting kinetic and biochemical information of time-dependent cellular processes. The technique allows the elucidation of the concentration profiles as well as the pure spectra of the components involved. Initially, we used Ordinary Differential Equations (ODE) to simulate drug uptake and 2 responses, which were employed to simulate perturbations to experimental control spectra, creating a dataset containing 36 simulated Raman spectra. Four different scenarios governing the drug exposure-response were evaluated: an undetectable disruption (e.g. radiation), a detectable disruption (e.g. a drug) and disruption with a signal significantly larger than the biological changes induced (e.g. a resonant drug), as well as simultaneous and asynchronous responses. Subsequently, data acquired from the exposure of a pulmonary adenocarcinoma cell line (A549) to Doxorubicin was analysed. The results indicate that MCR-ALS can independently identify and isolate both the spectra of the drug and the cell

responses under the different scenarios. The predicted concentrations map out the drug uptake and cellular response curves. The technique shows great potential to investigate non-linear kinetics and modes of action. Advantages and limitations of the technique are discussed, providing guidelines for future analysis strategies.

## Introduction

The EU and US (EU Directive-2010/63/EU and US Public Law 106-545, 2010, 106th Congress) have put in place legislation for the protection of animals used for scientific research and fostered research on the development of alternative methods for toxicity and safety testing to address legislative requirements. In this context, Raman microspectroscopy is a promising technique for the *in vitro* investigation of cellular drug uptake and modes of action.[1] The lateral resolution of confocal Raman microspectroscopy, typically in the hundreds of nanometre to micrometre range, allows spatial characterisation of organelles in cells[2]. Besides, as water is a weak Raman scatterer, the method is ideal for use in the living environment of cells[3]. By monitoring the spectra of the different cellular regions (e.g. cytoplasm, nucleus and nucleolus) after drug inoculation, a large amount of molecular information of the cell can be obtained, including the presence of the drug as well as metabolic changes induced. [4,5] Raman spectroscopy has been shown to be very efficient in monitoring the effect of the drug over time, which is essential to understand the drug mode (i.e. molecular targets of the drug) and mechanism of action (i.e. phenotypic changes induced by the drug).[6] By investigating the evolution of the Raman bands of the drug, the drug metabolites and the rest of the components, one can establish the composition ratio of the drug and the molecules which are affected by the drug. [7–10]

Different biomolecules (i.e. proteins, carbohydrates, nucleic acids and lipids), as well as drugs, can present highly overlapped Raman signatures and thus, the spectra can be used for extracting information about the drug uptake and the phenotypic responses induced by the drug. However, Raman data sets are complex and should be data mined using multivariate methods to extract biologically meaningful information.[11]

Although in most of the cases the cell functional responses as well as drug metabolism and uptake are changing in a non-linear fashion with time, linear algorithms, such as Partial Least Squares Regression (PLSR),[12] are commonly used to model these effects.[13–15] This implies building a model in which the predicted Y is the time value, and then studying the contribution of the different variables (either negative or positive) to the linear model, as manifest in the regression vector. As an alternative, orthogonal partial least squares regression (OPLSR) is an

extension of PLS which applies an orthogonal correction filter to extract spectral variations important to the prediction from the variations orthogonal to the time.[16] Although the performance of OPLSR models is generally largely equivalent to that provided by PLSR, OPLSR models gain in terms of interpretability of the variables, all the variance associated with Y being represented in the first latent variable (LV). Other methods to investigate these systems include two-dimensional[17] correlation analysis and Principal Component Analysis[18].

Two main problems are anticipated when using PLSR based techniques for the investigation of the temporal evolution of responses associated with drug modes and mechanisms of action. Firstly, while the Raman signal is linear with the concentration of Raman active biochemical compounds, both drug uptake and cellular responses follow diverse non-linear kinetic patterns [19]. Secondly, the regression vector cannot decompose the spectral variations which occur at different rates, being unable to differentiate spectral changes associated with the drug from variations associated with the subsequent cellular response. This restricts the use of PLSR to different time ranges in order to capture different linear regions of the drug uptake and response.[20]

Multivariate Curve Resolution - Alternating Least Squares (MCR-ALS) is used for resolving multiple component responses of unresolved mixtures, resolving the bilinear model by an alternative least squares iterative optimisation.[21] The bilinear model is defined as:

$$\mathbf{D} = \mathbf{C} \mathbf{S}^T + \mathbf{E} \quad (1) \quad (\text{Equation 1})$$

In which  $\mathbf{D}$  ( $i \times j$ ) is the matrix of spectral data of  $i$  samples and  $j$  variables, and  $\mathbf{C}$  ( $i \times n$ ) is the concentration matrix containing the concentrations of the  $N$  components of the system for each spectrum.  $\mathbf{S}^T$  ( $n \times j$ ) is the transpose of the spectra of the pure components and  $\mathbf{E}$  ( $i \times j$ ) is the residual matrix. When a temporal evolution is considered, MCR-ALS extracts: i) the concentration  $\mathbf{C}$  matrix of the components over  $i$  time points and ii) the pure spectrum of each component on the matrix  $\mathbf{S}$ . The former matrix is useful for studying the kinetics of the process and the latter to assign the bimolecular aetiology of each component. This technique is extremely useful in monitoring changes in complex and potentially non-linear temporal evolutions of multicomponent systems such as biological ones. It is also based on a bilinear model of Raman intensities and concentration of components

in a sample. However, MCR-ALS decomposes the spectral data matrix for a selected number of pure component spectra and concentration vectors. The concentration profiles can follow non-linear dependences with time, as defined by multiple kinetic equations.

In this work, using simulated and real data, we evaluate the use of MCR-ALS to decompose the Raman spectra of cells upon inoculation of a drug, to establish a kinetic model of the drug uptake, and cellular responses. First, we used Ordinary differential Equations (ODE) to simulate drug uptake and 2 subsequent response curves, which were added as time dependent perturbations to control cellular spectra, creating a dataset containing 36 simulated Raman spectra. Different scenarios were considered: The first represented the inclusion of an external perturbation which does not have a detectable Raman signal, simulating for example the effect of radiation or a drug or toxic with concentration below the limit of detection. The second scenario was simulated considering a drug or toxic with a specific Raman signature, with an intensity comparable to the cell component bands. The last scenario simulated the inoculation of a resonant drug with a Raman signal which is significantly stronger than the cell Raman signatures, potentially resonantly enhanced. Regarding the phenotypic changes induced by the presence of the drug, both simultaneous and asynchronous responses were analysed. MCR-ALS was applied considering different constraints, including non-negativity in the concentrations and spectra, and the results obtained were compared to those obtained by OPLSR. The results indicate that MCR-ALS can isolate and capture both the spectra of the drug and the cell responses under the different scenarios, in comparison with OPLSR, which only provided information about the mean trends of the spectral changes with time. Finally, a real dataset was studied, derived from the *in vitro* subcellular analysis of the time dependent uptake and localisation of the drug doxorubicin (DOX) in human lung cancer cells (A549)[20]. The predicted concentrations map out the drug uptake and cellular response curves. The technique shows great potential to investigate non-linear kinetics and modes of action.

## Materials and Methods

Here, we provide a summary of the materials and methods used for the experimental data. Further information can be accessed in the original publications[17,20].

### Cell culture and inoculation

A549 cells were cultured in DMEM-F12 with 10% fetal bovine serum (FBS) and Calu-1 cells in RPMI with 10% FBS, both at 37°C in a humidified atmosphere containing 5% CO<sub>2</sub> and cells were split every two days to maintain ~60% confluency.

Cells (~ 1x 10<sup>4</sup>/window) were seeded and incubated on CaF<sub>2</sub> windows (Crystan Ltd, UK) for 24 h for both control and exposure to DOX. The medium was then removed and samples were rinsed twice with sterile PBS and covered with DOX at the corresponding IC<sub>50</sub> inhibitory concentration, adjusted to the cell number[20]. After each incubation period, 2, 6, 12, 24, 48 and 72 h, cells were washed twice with sterile PBS and fixed in formalin (4%, 15 min).[22]

### Raman Spectroscopy

A Horiba Jobin-Yvon LabRAM HR800 spectrometer with a 785 nm, 300 mW diode laser as source, Peltier cooled 16-bit CCD, 300 lines/mm grating and 100 µm confocal hole, was used for this work. Spectra were acquired in the range from 400 cm<sup>-1</sup> to 1800 cm<sup>-1</sup> using a x100 objective (LCPlanN, Olympus), in dry conditions, for 30s two times, from three cell locations: cytoplasm, nucleus and nucleolus, visible under white light illumination. A final data set of 30 spectra per cell location for each time point, 2, 6, 12, 24, 48 and 72 h was produced after DOX exposure and for control cells, for each cell line, amounting to a total of over 210 cells per cell line, corresponding to a total data set of 1260 spectra. For the purpose of this study, data from the nucleolus only were considered.

### Data Simulation

Experimental control spectra were derived from the study of Farhane et al.[20] , measured as described above, and the 30 nucleolar spectra averaged. Based on the results of the PLSR analysis of the original study, simulated early (Response 1) and late (Response 2) spectral response profiles were generated in the Horiba LabSpec 5 software, as described in Byrne et al.[17] The temporal evolution of the respective contributions of the simulated responses were derived according to the rate equations described in the “Results Section”. Simulated datasets were created by adding weighted sums of the responses to the control spectra using the Matlab 2019a platform (Mathworks, Natick, MA, USA).

### Data Analysis

Data Treatment and data analysis is available at the *Zenodo* repository (10.5281/zenodo.3339285). Data analysis was performed using MATLAB 2019a from Mathworks (Natick, MA, USA). MCR-ALS was performed using the Graphical user interface available at <https://mcrals.wordpress.com/>. [23,24] Non negativity constraints were used for both spectra and concentrations. Considering the absence of maxima and minima in the concentration profiles, a unimodality constraint was also applied to the concentration. Pure spectra were normalised using the total sum (i.e. Each spectrum was divided by the total sum of the spectra). This step ensured that all the pure spectra computed were in a similar range. A maximum of 50 iterations were considered, employing a tolerance sigma value of 0.1. The initial estimations of the concentrations were obtained using those column profiles from the original data matrix which were numerically most different. The number of components tested varied from 2 to 3, depending on the Single Value Decomposition analysis performed prior to the MCR-ALS. Two kinds of MCR-ALS models were considered: without equality constraints and with equality constraints involving one component. In the latter case, the pure spectrum was forced to be the initial spectrum (InC) and the concentration was fixed to 1 for all the time points. Firstly, MCR-ALS was used without any equality constraint. If the model provided results with rotational ambiguity, the constraint was introduced.

OPLSR models were calculated using the PLS\_Toolbox 8.6 from Eigenvector Res. Inc. (Manson, WA, USA). Standard Normal Variate (SNV) and mean centring was used for spectral pre-processing. Model complexity (i.e.



number of LVs) was selected as the smallest number for which a decrease in the RMSECV between consecutive LVs was less than 3%.

## Results

### Simulation of spectral responses of cells under different scenarios

Simulations were based on previous works by Farhane et al.[20] and Byrne et al[17], which identified spectral responses of A549 cells to the inoculation of DOX over time, in the subcellular regions of nucleolus, nucleus and cytoplasm. Considering only the example of the nucleolus, the studies indicated that the interaction of DOX induced two perturbations to the spectra. The first response (Mode of action) was synchronous with the drug uptake and was caused by the binding of the drug to the RNA and DNA, resulting in a decrease of phosphate bands, including the  $785\text{ cm}^{-1}$  band (associated with RNA and DNA O-P-O stretching), the  $1095\text{ cm}^{-1}$  band (assigned to the DNA  $\text{PO}_2^-$  symmetric stretching), and the  $1683\text{ cm}^{-1}$  (protein Amide I). The second response (Mechanism of action) was subsequent to the drug uptake and was related to the phenotypic response of the cell to the drug, including an increased intensity of the RNA-sugar phosphate band at  $1047\text{ cm}^{-1}$ , amide III band at  $1271\text{ cm}^{-1}$  and  $\text{CH}_2$  deformation band at  $1444\text{ cm}^{-1}$ . Figure 1 shows the simulated spectral responses (Response 1 and Response 2) caused by the drug, along with the spectrum of the drug itself.

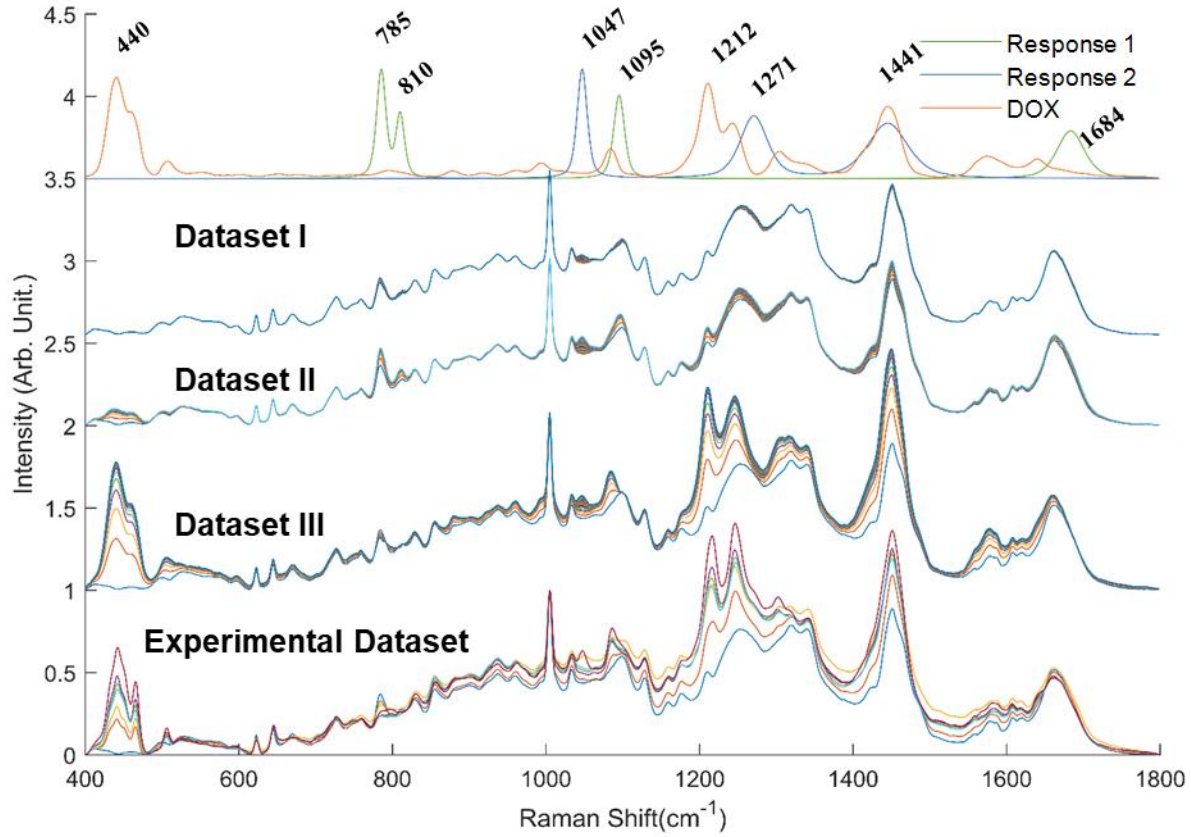


Figure 1. (Colour) Spectral profiles of the initial component (IC) simulated responses and experimental spectra of DOX (Top). Simulated and real datasets employed on the study. Dataset I: Undetectable disruption, subtractive Response 1 and additive Response 2. Dataset II: Detectable disruption, additive Response 1 and additive Response 2. Dataset III: Enhanced response (e.g. DOX), subtractive Response 1 and additive Response 2.

The effects of drug uptake after inoculation on a cell, described by the intensity of Responses 1 and 2, can be simulated considering ordinary differential equations[17].  $N_b$ , the number of DOX molecules bound to DNA and  $N_r$ , the intensity of the subsequent cellular response as a function of time can be expressed as:

$$\frac{dN_b}{dt} = (N_{recp} - N_b)K_{up}D \quad (2)$$

$$\frac{dN_r}{dt} = (N_{resp} - N_r)K_{resp}N_b \quad (3)$$

Where  $K_{up}$  and  $K_{resp}$  describe the drug uptake and response rates, respectively ( $2.5 \times 10^{-4}$  and  $2.5 \times 10^{-6} \text{ h}^{-1}$  in this study),  $D$  is the drug dose,  $N_{recp}$  is the number of receptors in the nucleolus and  $N_{resp}$  is a parameter to limit

the number of responses. To simulate the measured cellular responses, the spectral responses, weighted according to their temporal evolution, can be added to a measured, control dataset.

In this study, the use of MCR-ALS to different scenarios in which the phenotype of a cell is disrupted by an external agent was evaluated in simulated spectral data sets using the profiles calculated from equations (1) and (2) (see Figure 2a), and considering extrinsic perturbations with and without identifiable Raman spectral signatures as well as additive and subtractive spectral contributions.

*Dataset I. Undetectable perturbation, subtractive response 1 and additive response 2.*

This dataset simulates the effect of a perturbing factor which does not have its own a characteristic Raman spectrum, as is the case, for example, for ionising or non-ionising radiation, or a drug or toxicant which does not have a strong Raman signature at the dose applied [25]. The presence of this perturbation provokes a synchronous alteration of the cellular composition (i.e. Response 1, e.g. DNA damage) that triggers a subsequent molecular response of the cell (Response 2). Such a response can be simulated according to equation 4:

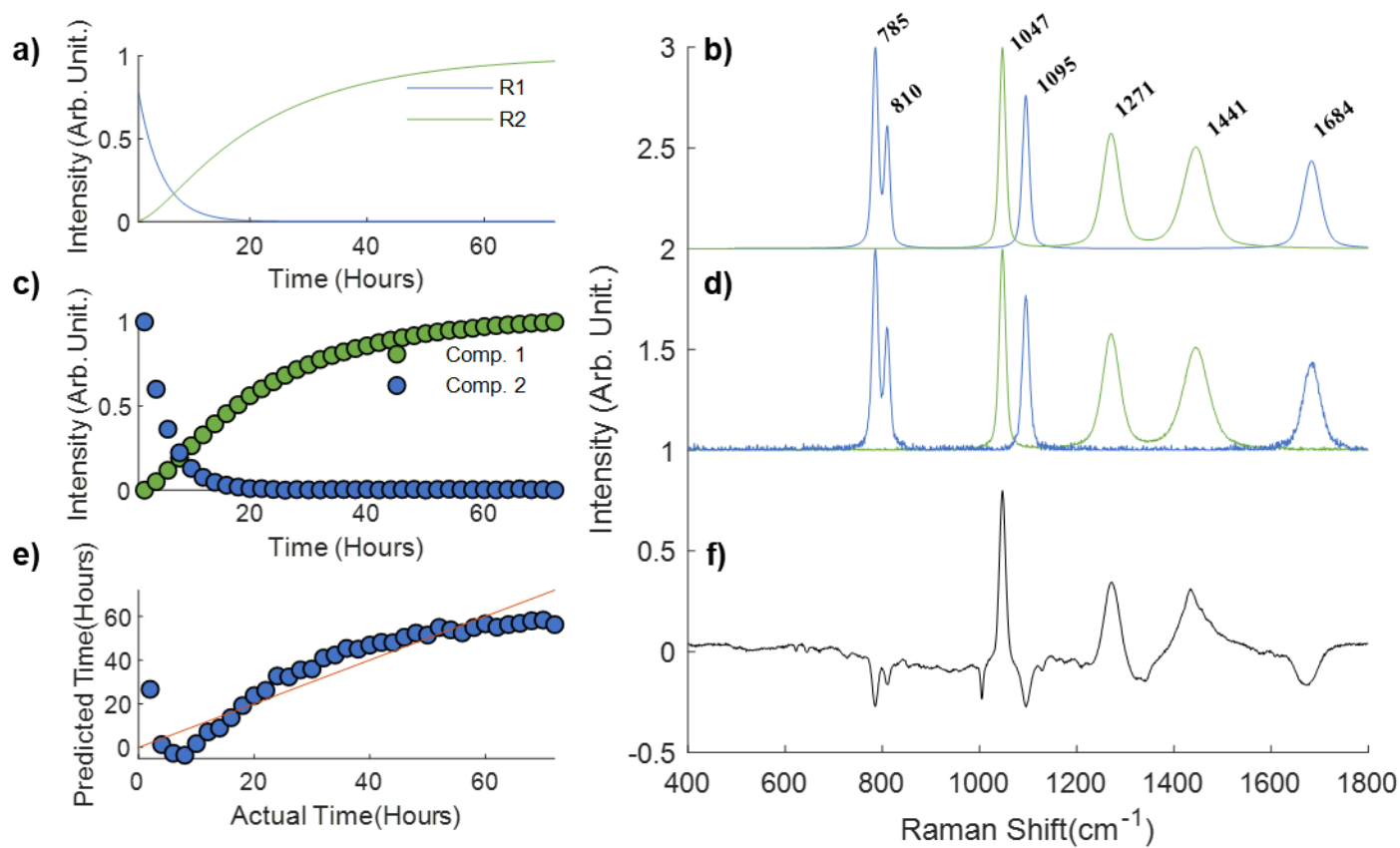
$$\mathbf{S}_1(t) = \mathbf{IC} - \mathbf{S}_{R1}\mathbf{I}_{R1}(t) + \mathbf{S}_{R2}\mathbf{I}_{R2}(t) + \mathbf{N} \quad (4)$$

Where IC is the initial component (i.e. Raman spectrum of the cell without the drug),  $\mathbf{S}_{R1}$  and  $\mathbf{S}_{R2}$  are the simulated spectra of the Responses 1 and 2, respectively and  $\mathbf{I}_{R1}(t)$  and  $\mathbf{I}_{R2}(t)$  are the simulated intensity profiles of Responses 1 and 2, respectively. N is Gaussian noise, computed as the 0.1 % of the length (maximum of the dataset minus minimum of the dataset) of the spectra ).

The simulated data set is depicted in Figure 1, showing the decrease of RNA band intensities associated with the simulated Response 1 and the increase of the intensity of protein bands due to the simulated subsequent Response 2. Figures 2a and b show the intensity and spectral profiles of the responses used for the simulation. The spectral profiles obtained from the MCR-ALS analysis (Figure 2d) were highly correlated with the simulated spectra. MCR-ALS components 1 and 2 were assigned to the simulated responses 2 and 1, respectively, and the Pearson correlation coefficients between the reference and predicted spectra were higher than 0.99. The variations of the concentration of these components as a function of time are depicted in Figure 2c. The concentration of MCR-

ALS component 2 showed a monotonic decrease, reaching a plateau after 20 h, in agreement with the simulated concentration of Response 1. In contrast, MCR-ALS component 1 showed an increase in concentration, highly correlated with the simulated concentration of Response 2. Thus, MCR-ALS was able to extract both the spectral responses and the time dependant profiles of both responses. Results from the analysis of the simulated data set by OPLSR are shown in Figures 2e and 2f. The first OPLSR loading, which represents the spectral variance associated with time, showed negative bands at 785, 811, 1099 and 1670  $\text{cm}^{-1}$  (i.e. Response 1) and positive bands at 1047, 1275 and 1436  $\text{cm}^{-1}$  (i.e. Response 2). Cross validated predicted time values showed a significant correlation between the simulated spectra and the time (see Figure 2e). However, a basic assumption of PLSR is that the association between the changes in the spectra and the response (i.e. time) is linear. The simulated data set included non-linear responses and so, the generalisation performance of PLSR is limited and the model may not be adequate as shown by the non-uniform distribution of the residuals of the  $y$  predicted values. In contrast, MCR-ALS provides a more accurate description of the non-linear relationship between the two components with time. Asynchronous subtractive and additive responses were represented as decreasing and increasing signatures

on the concentration profiles, respectively.



*Figure 2. (Colour) MCR-ALS and OPLSR analysis of Dataset I (Undetectable disruption, subtractive Response 1 and additive Response 2). Simulated time evolution (a) and simulated spectral signatures (b) employed on the creation of Dataset I, simulating the spectral signatures of the disruption (Response 1) and subsequent cellular response (Response 2). Concentration (c) and spectral (d) profiles determined by the MCR-ALS. Actual time values versus predicted (by cross validation) time values (e) and first loading vector of the OPLSR (f). Concentration profiles on a) and c) have been normalised to the range [0 1]. Spectra in b) and d) have been shifted on the y axis for clarity.*

Dataset II. Detectable disruption, additive response 1 and additive response 2.

Dataset II simulates the effect of a Raman active perturbation factor, such as a drug or particle with a concentration above the limit of detection. Its subcellular interaction provokes a simultaneous Response 1, which represents the binding of the molecule or particle to the molecules of the cell and a subsequent phenotypic response of the cell (Response 2). Dataset II was simulated according to equation 5:

$$\mathbf{S2(t)} = \mathbf{IC} + \mathbf{S_{Drug}I_{Drug}(t)} + \mathbf{S_{R1}I_{R1}(t)} + \mathbf{S_{R2}I_{R2}(t)} + \mathbf{N} \quad (5)$$

in which  $\mathbf{S_{Drug}}$  and  $\mathbf{I_{Drug}}$  are the spectra and concentration profile of the perturbing agent, in this case the drug spectrum and its dynamic uptake, respectively. Here, we assumed a negligible concentration of free drug compared to the bound drug, and that the drug uptake and Response 1 occur simultaneously. Also, in this case we also changed the sign of Response 1 to explore the use of MCR-ALS under additive conditions (i.e. the perturbative interaction resulted in an increase in the contribution of constituent cellular components). A visual inspection of the spectra of dataset II (Figure 1) revealed spectral variations associated with Responses 1 and 2, as well as characteristic drug bands.

The accuracy of the MCR-ALS decomposition is demonstrated by the high degree of correlation observed between the simulated spectral profiles (Figure 3d) and the spectra of the components obtained by MCR-ALS (Figure 3b). Raman bands of the second MCR-ALS component were assigned to Response 2 and Raman bands of the first MCR-ALS component are consistent with a combination of the spectra of the drug and Response 1. As expected, MCR-ALS was not able to resolve the simulated drug internalisation and cellular response as two independent components, because both processes are synchronous. If they had different kinetics, it would have been possible to use the known DOX spectra as a pure spectrum to help resolve the system. Nonetheless, the MCR-ALS estimated concentration profiles obtained, depicted in Figure 3c, are highly correlated with the intensity profiles employed for the simulation (Figure 3a). In this case, the profile corresponding to the first MCR-ALS component, which represented the simultaneous uptake of drug and cell response increased over time.

Analysis by OPLSR (see Figure 3e and f) revealed again the lack of the linearity of the spectra and time. Although the first loading appropriately indicated an increase of DOX, Response 1 and Response 2 bands, MCR-ALS provided a more accurate description of the different non-linear spectral variations in the dataset over time.

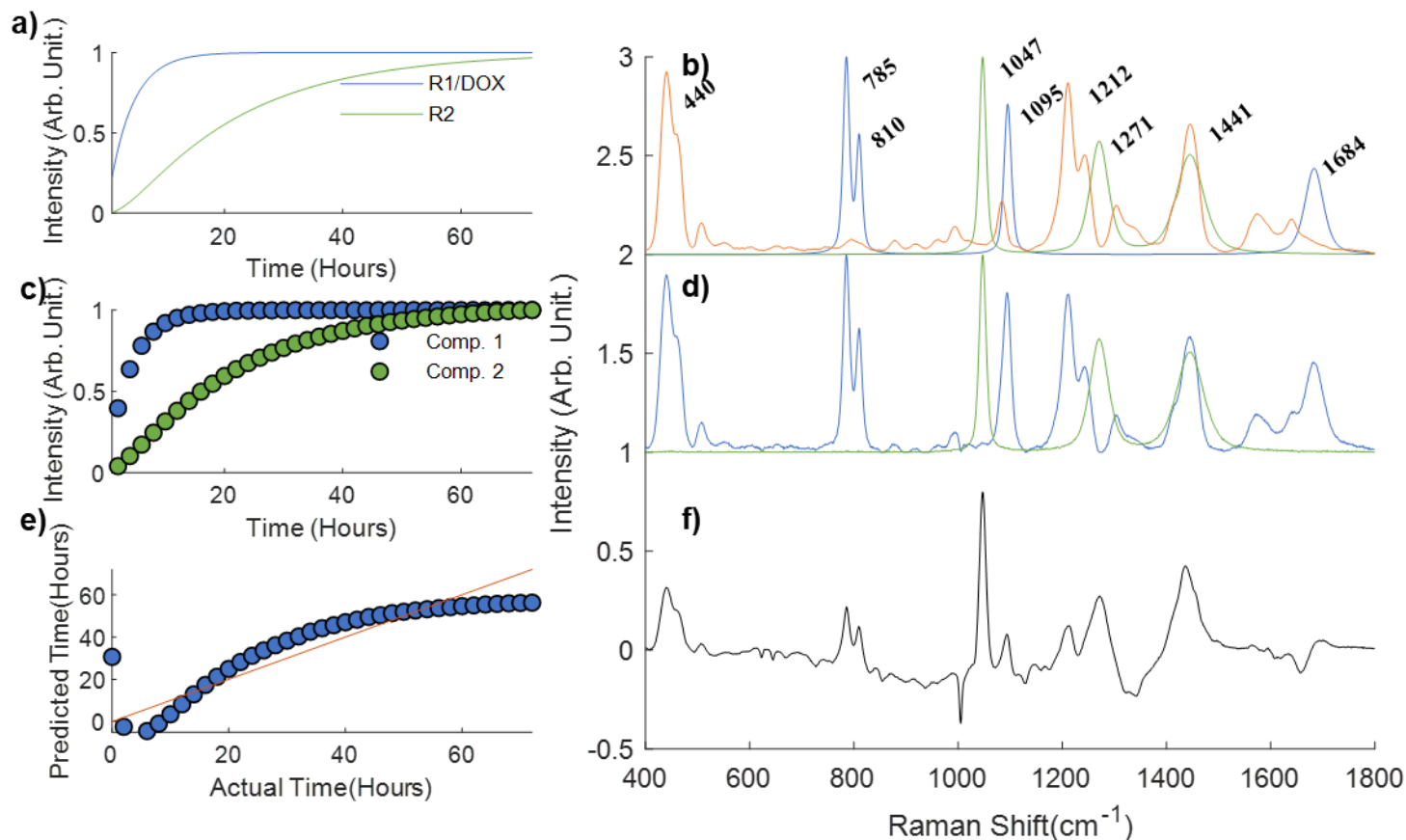


Figure 3. (Colour) MCR-ALS and OPLSR analysis of Dataset II (Detectable disruption, additive Response 1 and additive Response 2). Simulated time evolution (a) and simulated spectral signatures (b) employed on the creation of Dataset II, simulating the spectral signatures of the disruption (e.g. a drug) the bonding of the disruption (Response 1) and subsequent cellular responses (Response 2). Concentration (c) and spectral (d) profiles determined by MCR-ALS. Actual time values versus cross validated predicted time values (e) and first loading vector of the OPLSR (f). Concentration profiles on a) and c) have been normalised to the range [0 1]. Spectra on b) and d) have been shifted on the y axis for clarity.

#### Dataset III. Resonant Drug, subtractive response 1 and additive response 2.

Data III simulates specifically the *in vitro* inoculation of A549 cells with DOX. In this case, the perturbing effect is a resonant drug and therefore its signal is comparable or significantly more intense than the signal associated

with the drug bonding (Response 1) and subsequent cellular metabolic reaction (Response 2). The representative, Dataset III can be simulated according to:

$$S_3(t) = IC + F S_{Drug} I_{Drug}(t) - S_{R1} I_{R1}(t) + S_{R2} I_{R2}(t) + N \quad (6)$$

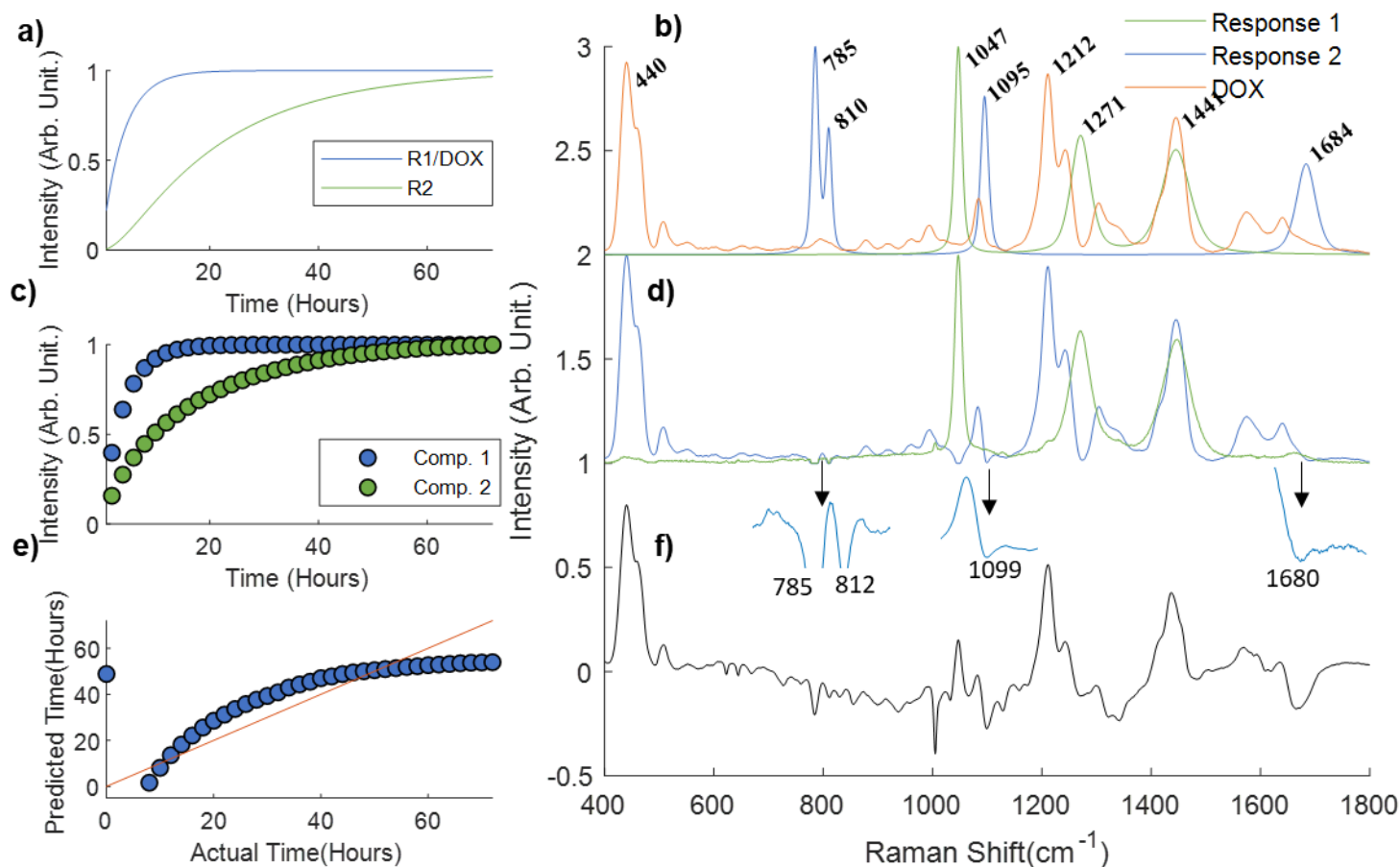
where F is a factor which increases the signal of the drug relatively to the signal of Response 1 and Response 2. In this study F=8 was selected to match the values observed in the real dataset. Here, Response 1 was negative, again to simulate the experimentally observed spectral signature of the drug bonding to RNA/DNA in the nucleolus of the cell.

Three components were considered for the MCR-ALS analysis of this dataset, the first (i.e. component 3) being the initial cellular control spectrum, which was time independent. Spectral bands of MCR-ALS component 2 (Figure 4b) were seen to be highly correlated with Response 2 ( $R^2 = 0.9935$ ), although the spectrum was slightly contaminated with bands from DOX, such as those at 1620 and 420  $\text{cm}^{-1}$  and control cellular components such as the Phenylalanine band at 1005  $\text{cm}^{-1}$ . The concentration profile of this component also followed that of Response 2 (Figure 4b). A visual inspection of the first MCR-ALS component revealed a high correlation with the DOX spectrum ( $R^2 = 0.9963$ ). Bands of Response 1 should also be present in this component, as they were simulated to occur simultaneously. However, in contrast to Dataset II, in this case they should produce a negative contribution, as in the case of Dataset I, for which the concentration profile indicated a decrease of the component. In the specific scenario considered here, Component 2 should represent the increase of DOX bands and the simultaneous decrease of Response 1 bands.

For dataset III, the concentration profile indicated a signal increase, similar to the simulated DOX uptake (see Figure 4c). A close inspection of the spectra of first MCR-ALS component confirms the presence of bands with a negative contribution at the Response 1 positions (insets below Figure 4d). Bands at 780 and 810  $\text{cm}^{-1}$  were flattened to the 0 value and bands at 1095 and 1685  $\text{cm}^{-1}$  were found as negative shoulders of DOX bands. In summary, Component 1 was a combination of a large contribution of the DOX spectrum minus a small contribution of Response 1, correctly reproducing the information introduced in the simulation. From these observations, it can be established that, in the case of a resonant drug, spectral signatures of the binding can be



masked by the large contribution of the drug, and that subtractive contributions will be represented as nodes in the spectra. For comparison, the first loading vector of the OPLSR (Figure 4f) shows a combination of the three sources of variation, dominated by the DOX bands, but also reveals small positive contributions of Response 2 and negative contributions of Response 1.



*Figure 4. (Colour) MCR-ALS and OPLSR analysis of Dataset III (Detectable disruption, subtractive Response 1 and additive Response 2). Simulated time evolution (a) and simulated spectral signatures (b) employed on the creation of Dataset III, simulating the spectral signatures of the perturbation (e.g. a resonant drug) the bonding of the drug (Response 1) and subsequent cellular responses (Response 2). Concentration (c) and spectral (d) profiles determined by MCR-ALS. Actual time values versus predicted (by cross validation) time values (e) and first loading vector of the OPLSR (f). Insets represent enlargement of the Component 2 spectral bands. Concentration profiles on a) and c) have been normalised to the range [0 1]. Spectra on d) and b) have been shifted on the y axis for clarity.*

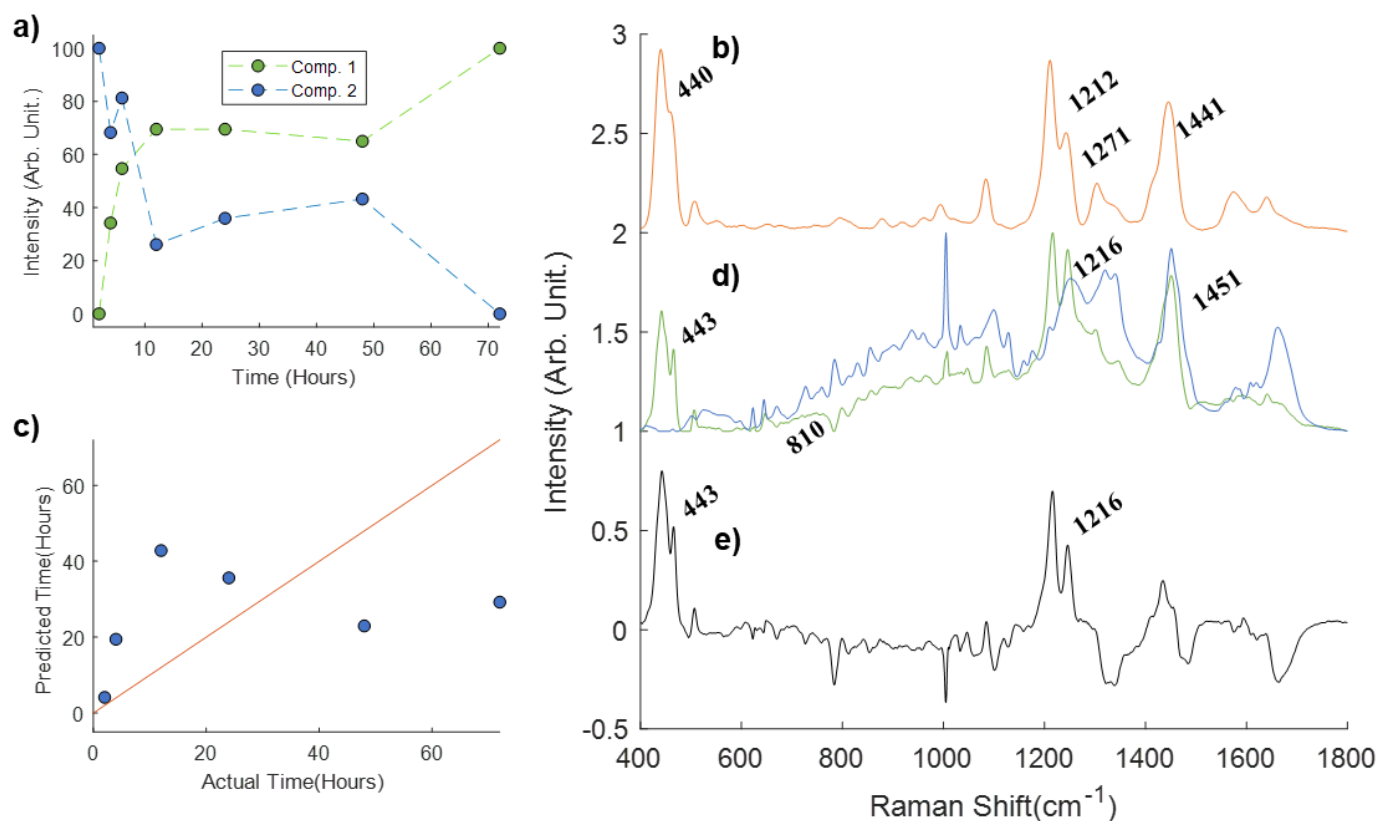
### Experimental DOX Dataset

The experimental DOX dataset, previously described by Farhane et al.[18], was analysed using MCR-ALS and OPLSR under the conditions described on the “Data Analysis” section. Figure 1 depicts the Raman dataset, which shows similarities with Dataset III. For MCR-ALS, two components were used, as a third component provided ambiguous results (see Electronic Supplementary Material, SM1). The pure spectrum of MCR-ALS Component 2 (Figure 5d, blue) shows typical bands of the original cell, whereas the spectra of MCR-ALS Component 1 (green) shows signatures of DOX bands and decreasing bands of DNA at 782, 812 and 1680  $\text{cm}^{-1}$ . The temporal evolution of the spectral profile (Figure 5a) indicates an increase of these bands, evidencing an increase of DOX content and a decrease of DNA bands, which reaches a plateau after 16 h. This indicates that the real data follows the simulated Response 1 of Dataset III. The OPLSR component loading (Figure 5d) also shows similar bands, but it does not provide any kinetic information about the spectral evolution over time.

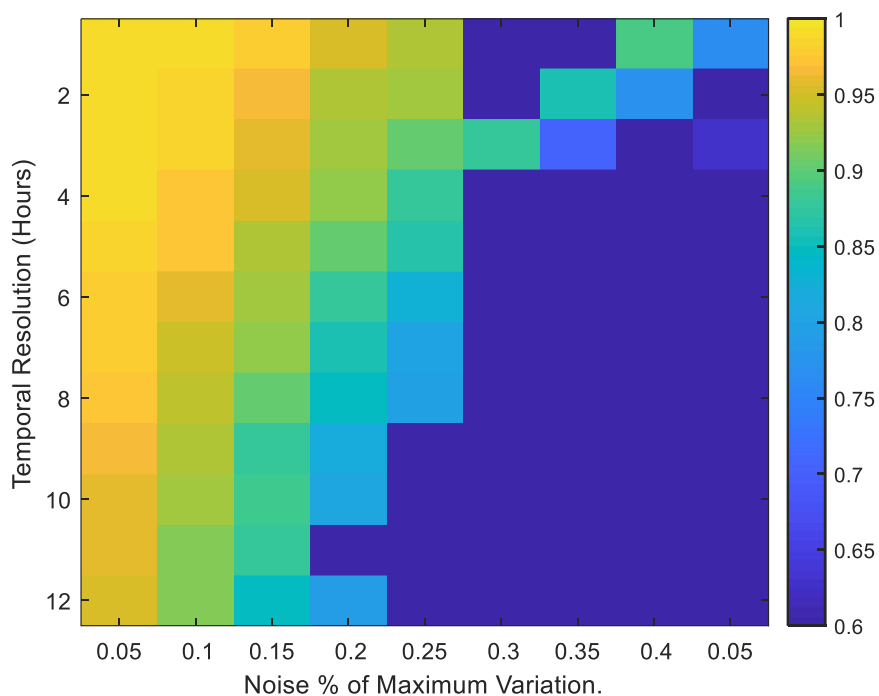
The ambiguity found when three components were employed confirms that the secondary cellular response, expected to be similar to the simulated Response 2, was not identified in the real dataset. Nevertheless, it has to be noted that the real dataset included only spectra collected at 7 time points, and this limited temporal resolution hampered the resolution of complex cellular signatures. In order to study this, a further simulation experiment with the Dataset III was performed in which MCR-ALS models from simulated data with different levels of noise and number of time points were simulated. Figure 6 shows the similarity between the Response 2 spectral signature and the MCR-ALS component calculated as the correlation coefficient, over different temporal resolutions and noise levels. Results showed that, regardless the noise level, the quality of the resolved concentration and spectral profiles improves with the number of time points.

In summary, the results indicate that MCR-ALS outperforms the OPLSR method in extracting kinetic information as well as spectral markers of the different processes involved. On the other hand, in a recent study using the same experimental data and simulation approach, the use of a two dimensional correlation analysis [17] method was also able to determine that bands associated with Response 1 and Response 2 are asynchronous (i.e. they occurred at different rates). It did it also for the real dataset, while in this study MCR-ALS was not able to detect the

Response 2, probably because the poor spatial resolution of the experimental data. However, compared to 2D correlation, MCR-ALS was able to provide the concentration profiles over time, which are extremely useful for the study of the kinetics involved on the drug uptake and cellular response.



*Figure 5. (Colour) MCR-ALS and OPLSR analysis of the Real Dataset (A549 cells exposed to DOX). Reference spectra of DOX (b) Concentration (a) and spectral (d) profiles determined by the MCR-ALS. Actual time values versus predicted by CV time values (c) and first loading vector of the OPLSR (e). Spectra on b) and d) have been shifted on the y axis for clarity.*



*Figure 6. (Colour) Correlation coefficient of the spectral signature of Response 2 and the pure spectra extracted by the MCR-ALS for datasets simulated using different amount of noise and temporal resolution.*

## Conclusions

Results indicated that MCR-ALS is a usefully method for investigating the temporal evolution of drug uptake and cellular responses in cells in vitro. It outperforms OPLSR, as it provides kinetic information from different components that can evolve according to different non-linear time profiles. Dataset I indicated that additive and subtractive perturbations of the constituent cellular components can be identified in the concentration profiles and pure spectra profiles. Dataset II showed that simultaneous changes in molecules appear integrated as a single component, the use of molecular information (e.g. prior knowledge about the location of drug bands) being necessary to differentiate contributions of different species. Dataset III indicated that, if a component represents both, additive and subtractive contributions, the spectrum calculated by MCR-ALS will contain positive bands and nodes respectively.

MCR-ALS applied to real data from DOX inoculations in A549 cells extracted a component representing the drug uptake as well as the binding to DNA and RNA molecules. The dataset was limited in terms of temporal resolution, which hampered the extraction of the subsequent, longer term cellular responses. In summary, the application of MCR-ALS to decompose time series of Raman spectra after drug inoculation extracted significant information about changes on cell composition induced by the presence of the drug. This could be applied in the investigation of drug kinetics and modes of action and could be regarded as a revolutionary technique in preclinical screening to improving drug effectivity and investigate drug resistance.

## Acknowledgments

**DPG** acknowledges support from the European Research Council MSCA grant (Spectro-metrics, 020-MSCA-IF-2017 Project ID:796287). **GQ** acknowledges support from the *Agencia Estatal de Investigación* (AEI) and the *Fondo Europeo de Desarrollo Regional* (FEDER) (CTQ2016-79561-P).

## References

- [1] B. Kann, H.L. Offerhaus, M. Windbergs, C. Otto, Raman microscopy for cellular investigations — From single cell imaging to drug carrier uptake visualization, *Adv. Drug Deliv. Rev.* 89 (2015) 71–90. doi:10.1016/j.addr.2015.02.006.
- [2] E. Efeoglu, M. Keating, J. McIntyre, A. Casey, H.J. Byrne, Determination of nanoparticle localisation within subcellular organelles in vitro using Raman spectroscopy, *Anal. Methods.* 7 (2015) 10000–10017. doi:10.1039/C5AY02661J.
- [3] M. Jermyn, K. Mok, J. Mercier, J. Desroches, J. Pichette, K. Saint-Arnaud, L. Bernstein, M.-C. Guiot, K. Petrecca, F. Leblond, Intraoperative brain cancer detection with Raman spectroscopy in humans, *Sci. Transl. Med.* 7 (2015) 274ra19-274ra19. doi:10.1126/scitranslmed.aaa2384.
- [4] R. Smith, K. L. Wright, L. Ashton, Raman spectroscopy: an evolving technique for live cell studies, *Analyst.* 141 (2016) 3590–3600. doi:10.1039/C6AN00152A.
- [5] S.F. El-Mashtoly, H.K. Yosef, D. Petersen, L. Mavarani, A. Maghnouj, S. Hahn, C. Kötting, K. Gerwert, Label-Free Raman Spectroscopic Imaging Monitors the Integral Physiologically Relevant Drug Responses in Cancer Cells, *Anal. Chem.* 87 (2015) 7297–7304. doi:10.1021/acs.analchem.5b01431.
- [6] Z. Farhane, H. Nawaz, F. Bonnier, H.J. Byrne, In vitro label-free screening of chemotherapeutic drugs using Raman microspectroscopy: Towards a new paradigm of spectralomics, *J. Biophotonics.* 11 (2018) e201700258. doi:10.1002/jbio.201700258.
- [7] S.F. El-Mashtoly, D. Petersen, H.K. Yosef, A. Mosig, A. Reinacher-Schick, C. Kötting, K. Gerwert, Label-free imaging of drug distribution and metabolism in colon cancer cells by Raman microscopy, *The Analyst.* 139 (2014) 1155–1161. doi:10.1039/c3an01993d.
- [8] S.-S. Li, Q.-Y. Guan, G. Meng, X.-F. Chang, J.-W. Wei, P. Wang, B. Kang, J.-J. Xu, H.-Y. Chen, Revealing chemical processes and kinetics of drug action within single living cells via plasmonic Raman probes, *Sci. Rep.* 7 (2017) 2296. doi:10.1038/s41598-017-02510-9.

- [9] Y. Harada, P. Dai, Y. Yamaoka, M. Ogawa, H. Tanaka, K. Nosaka, K. Akaji, T. Takamatsu, Intracellular dynamics of topoisomerase I inhibitor, CPT-11, by slit-scanning confocal Raman microscopy, *Histochem. Cell Biol.* 132 (2009) 39–46. doi:10.1007/s00418-009-0594-0.
- [10] K. Aljakouch, T. Lehtonen, H.K. Yosef, M.K. Hammoud, W. Alsaidi, C. Kötting, C. Mügge, R. Kourist, S.F. El-Mashtoly, K. Gerwert, Raman Microspectroscopic Evidence for the Metabolism of a Tyrosine Kinase Inhibitor, Neratinib, in Cancer Cells, *Angew. Chem. Int. Ed Engl.* 57 (2018) 7250–7254. doi:10.1002/anie.201803394.
- [11] R. Gautam, S. Vanga, F. Ariese, S. Umapathy, Review of multidimensional data processing approaches for Raman and infrared spectroscopy, *Epj Tech. Instrum.* 2 (2015) 8. doi:10.1140/epjti/s40485-015-0018-6.
- [12] M.E. Keating, H. Nawaz, F. Bonnier, H.J. Byrne, Multivariate statistical methodologies applied in biomedical Raman spectroscopy: assessing the validity of partial least squares regression using simulated model datasets, *Analyst.* 140 (2015) 2482–2492. doi:10.1039/C4AN02167C.
- [13] H. Nawaz, F. Bonnier, P. Knief, O. Howe, F.M. Lyng, A.D. Meade, H.J. Byrne, Evaluation of the potential of Raman microspectroscopy for prediction of chemotherapeutic response to cisplatin in lung adenocarcinoma, *The Analyst.* 135 (2010) 3070. doi:10.1039/c0an00541j.
- [14] H. Nawaz, A. Garcia, A.D. Meade, F.M. Lyng, H.J. Byrne, Raman micro spectroscopy study of the interaction of vincristine with A549 cells supported by expression analysis of bcl-2 protein, *Analyst.* 138 (2013) 6177–6184. doi:10.1039/C3AN00975K.
- [15] H. Nawaz, F. Bonnier, A.D. Meade, F.M. Lyng, H.J. Byrne, Comparison of subcellular responses for the evaluation and prediction of the chemotherapeutic response to cisplatin in lung adenocarcinoma using Raman spectroscopy, *Analyst.* 136 (2011) 2450–2463. doi:10.1039/C1AN15104E.
- [16] E.K. Kemsley, H.S. Tapp, OPLS filtered data can be obtained directly from non-orthogonalized PLS1, *J. Chemom.* 23 (2009) 263–264. doi:10.1002/cem.1217.
- [17] H.J. Byrne, F. Bonnier, Z. Farhane, Two-dimensional correlation analysis of Raman microspectroscopy of subcellular interactions of drugs in vitro, *J. Biophotonics.* 12 (2019) e201800328. doi:10.1002/jbio.201800328.
- [18] Z. Farhane, F. Bonnier, M.A. Maher, J. Bryant, A. Casey, H.J. Byrne, Differentiating responses of lung cancer cell lines to Doxorubicin exposure: in vitro Raman micro spectroscopy, oxidative stress and bcl-2 protein expression, *J. Biophotonics.* 10 (2017) 151–165. doi:10.1002/jbio.201600019.
- [19] H.J. Byrne, M.A. Maher, Numerically modelling time and dose dependent cytotoxicity, *Comput. Toxicol.* 12 (2019) 100090. doi:10.1016/j.comtox.2019.100090.
- [20] Z. Farhane, F. Bonnier, H.J. Byrne, Monitoring doxorubicin cellular uptake and trafficking using in vitro Raman microspectroscopy: short and long time exposure effects on lung cancer cell lines, *Anal. Bioanal. Chem.* 409 (2017) 1333–1346. doi:10.1007/s00216-016-0065-0.
- [21] A. de Juan, J. Jaumot, R. Tauler, Multivariate Curve Resolution (MCR). Solving the mixture analysis problem, *Anal. Methods.* 6 (2014) 4964–4976. doi:10.1039/C4AY00571F.
- [22] Z. Farhane, F. Bonnier, O. Howe, A. Casey, H.J. Byrne, Doxorubicin kinetics and effects on lung cancer cell lines using in vitro Raman micro-spectroscopy: binding signatures, drug resistance and DNA repair, *J. Biophotonics.* 11 (2018) e201700060. doi:10.1002/jbio.201700060.
- [23] J. Jaumot, R. Gargallo, A. de Juan, R. Tauler, A graphical user-friendly interface for MCR-ALS: a new tool for multivariate curve resolution in MATLAB, *Chemom. Intell. Lab. Syst.* 76 (2005) 101–110. doi:10.1016/j.chemolab.2004.12.007.
- [24] J. Jaumot, A. de Juan, R. Tauler, MCR-ALS GUI 2.0: New features and applications, *Chemom. Intell. Lab. Syst.* 140 (2015) 1–12. doi:10.1016/j.chemolab.2014.10.003.
- [25] Z. Farhane, F. Bonnier, H.J. Byrne, An in vitro study of the interaction of the chemotherapeutic drug Actinomycin D with lung cancer cell lines using Raman micro-spectroscopy, *J. Biophotonics.* 11 (2018). doi:10.1002/jbio.201700112.